A ISOMORPHISM-INVARIANT RELATIONAL VARIABLE

We describe a desired property of a relational variable. This property ensures that the interpretation of relational variable is consistent across graph isomorphic connected components of any relational skeleton $\sigma \in \Sigma_{\mathcal{S}}$. Let $\mathcal{S} = \langle \mathbf{E}, \mathbf{R}, \mathbf{A} \rangle$ be a relational schema and $\sigma \in \Sigma_{\mathcal{S}}$ be an arbitrary instantiation of the schema. Let $\mathrm{CC}_a^\sigma = \langle \mathcal{V}_a, \mathcal{E}_a, \mathcal{L}_a \rangle$ and $\mathrm{CC}_b^\sigma = \langle \mathcal{V}_b, \mathcal{E}_b, \mathcal{L}_b \rangle$ where each item is labeled with its item class. Let $\mathbf{f}_{a,b}^\sigma$ be a set of mapping functions demonstrating $\mathrm{CC}_a^\sigma \cong_{a,b} \mathrm{CC}_b^\sigma$, that is, $\mathbf{f}_{a,b}^\sigma = \{f \mid \forall_{v \in \mathcal{V}_a} \mathcal{L}_a(v) = \mathcal{L}_b(f(v)) \wedge \forall_{u,v \in \mathcal{V}_a} (u, v) \in \mathcal{E}_a \Leftrightarrow (f(u), f(v)) \in \mathcal{E}_b \wedge f(a) = b\}$. Let U be a relational variable with domain $\sigma(I)$ where $I \in \mathbf{E} \cup \mathbf{R}$. If

$$\forall_{f \in \mathbf{f}_{a,b}^\sigma} \{f(i) \mid i.X \in U_a\} = \{j \mid j.X \in U_{f(a)}\}$$

for any $a, b \in \sigma(I)$ for any $\sigma \in \Sigma_{\mathcal{S}}$, then, U is said to be *isomorphism-invariant.*

B EXPERIMENTAL SETUP

We describe experimental settings with the language of relational causal model (Maier et al., 2013). For simplicity, we represent a relational skeleton as an undirected graph of entities. Hence, $(a, b) \in \sigma$ represents that two entities a and b are, in fact, connected to a common relationship item in the relational skeleton σ.

B.1 SIMPLE EXPERIMENTS

Relational Schema There are four entity classes A, B, C, and D where each associates with attribute class X, Y, S, and T, respectively. There are three relationship classes between A and B, A and C, and B and D. We used circle, square, rhombus, and triangle to refer A, B, C, and D.

Relational Skeleton We generate relational skeletons varying degrees of randomness from 0 to 1 where randomness of 0 is referred to as 'biased'. Given randomness $0 \leq p \leq 1$, we initially generate a fully biased relational skeleton σ, then randomize some of edges between $\sigma(A)$ and $\sigma(B)$ to acquire a relational skeleton of desired randomness p.

Let n be the number of entities for each of A and B (we set $n = 200$). There are $n/2$ entities for C and D, respectively. Hence, let $\sigma(A) = \{a_i\}_{i=1}^n$, $\sigma(B) = \{b_i\}_{i=1}^n$, $\sigma(C) = \{c_i\}_{i=1}^{n/2}$, and $\sigma(D) = \{d_i\}_{i=1}^{n/2}$. We connect C items to A items and D items to B items, $\{(a_i, c_i)\}_{i=1}^{n/2} \subset \sigma$ and $\{(b_i, d_i)\}_{i=1}^{n/2} \subset \sigma$. Then, we "initially" connect a_i and b_i, $\{(a_i, b_i)\}_{i=1}^{n} \subset \sigma$. That is, by design, an A item having (or not having) a C neighbor is connected to a B item having (or not having) a D neighbor. Given the randomness p, we randomly pick np B items and shuffle their A neighbors.

Relational Causal Model We use a linear Gaussian noise model with sum aggregators as follows:

$$\begin{aligned}
&\forall_{c \in \sigma(C)} \quad c.S = \mu + \epsilon_c \\
&\forall_{d \in \sigma(D)} \quad d.T = \mu + \epsilon_d \\
&\forall_{a \in \sigma(A)} \quad a.X = \sum_{c \in ne^\sigma(a) \cap \sigma(C)} c.S + \mu + \epsilon_a \\
&\forall_{b \in \sigma(B)} \quad b.Y = \sum_{a \in ne^\sigma(b) \cap \sigma(A)} \beta \cdot a.X + \\
&\qquad\qquad\qquad \sum_{d \in ne(b;\sigma) \cap \sigma(D)} d.T + \mu + \epsilon_b
\end{aligned}$$

where every ϵ is an independent Gaussian noise with zero mean and variance 0.1^2. We control the correlation between connected X and Y by adjusting β where $\beta = 0$ implies that X and Y are independently generated, or more precisely, $[A].X$ and $[A, R_{AB}, B].Y$ are independent. The pair of X and Y values generated from this model can be understood as a mixture of four bivariate normal distributions and μ controls the distance between distributions. When $\mu = 0$, all four distributions are centered at $(0, 0)$. Although, we described four distributions having the same mean as 'homogeneous', they have different variances.

We test unconditional independence between $[A].X$ and $[A, R_{AB}, B].Y$.

B.2 MORE COMPLICATED EXPERIMENTS

We only change how skeletons are generated. We use the same relational schema and relational causal model as shown above.

Relational Skeleton Similarly, we generate $n = 400$ items for A and B and $n/2$ items for C and D. Then, each $a_i \in \sigma(A)$ randomly chooses C neighbor(s) so that a_i has one C neighbor if $1 \leq i \leq \frac{n}{3}$, two neighbors if $\frac{n}{3} < i \leq \frac{2n}{3}$, and three neighbors if $\frac{2n}{3} < i \leq n$. Similarly, each $b_i \in \sigma(B)$ randomly chooses D neighbor(s). As shown in the previous setup, we initialize relational skeleton with biased relationships between A and B, that is, $\{(a_i, b_i)\}_{i=1}^n \subset \sigma$. Then, we randomize the connection based on randomness parameter. We further add n random connections between $\sigma(A)$ and $\sigma(B)$.

This setup yields more complicated structure than the previous setup since each of $\{a.X\}_{a \in \sigma(A)}$ and

$\{b.Y\}_{b\in\sigma(B)}$ is made of dependent observations and A and B are in many-to-many relationships.

B.3 CONDITIONAL TESTS

Relational Schema We have three entity classes A, B, and C, which associates with X, Y, and Z, respectively. There are binary relationship classes for each pair of entity classes, i.e., R_{AB}, R_{AC}, and R_{BC}. All cardinalities are '*many*', hence an entity can have *many* neighbors of the other entity class.

Relational Skeleton We control the maximum number of neighbors of the same kind. For example, an item of A can have at most k neighbors of B. In other words, $\forall_{a_i\in\sigma(A)} |ne^{\sigma}(a_i)\cap\sigma(B)| \leq k$. We similarly put restrictions between B and C and between A and C, as well.

We construct relational skeletons where relationships of all three classes (R_{AB}, R_{BC}, and R_{AC}) are correlated. To do so, we adopt the idea of latent space modeling. Given n, the number of entities per entity class, we generate n points in $[0,1]^2 \subset \mathbb{R}^2$ for each entity class. Let $\phi(\cdot)$ be the coordinate of an item. Let $\mathbf{D}^{AB}$ be a squared Euclidean distance where $\left(\mathbf{D}^{AB}\right)_{i,j} = \|\phi(a_i) - \phi(b_i)\|_2^2$. Then, a kernel matrix $\mathbf{K}^{AB}$ is $\left(\mathbf{K}^{AB}\right)_{i,j} = \exp\left(-\gamma\cdot\left(\mathbf{D}^{AB}\right)_{i,j}\right)$ where we chose $\gamma = 50$. By normalization, we get a probability matrix $\mathbf{P}^{AB} = \frac{\mathbf{K}^{AB}}{(\mathbf{1}^{\top}\cdot\mathbf{K}^{AB}\cdot\mathbf{1})}$ to (approximately) model $\Pr((a_i,b_j)\in\sigma) \propto 2\left(\mathbf{P}^{AB}\right)_{i,j}$. With this probability, we sample $nk/2$ edges to form a relational skeleton while satisfying maximum number of neighbors k. For example, if we limit an item of A can have three neighbors of B, then, there are, on average, 1.5 B neighbors for an item of A. Edges between A and C and between B and C are similarly obtained.

Relational Causal Model We consider three different models: two for conditional independence and one for conditional dependence. For testing null hypothesis, we randomly choose one of following two models:

$$\begin{aligned}
\forall_{a\in\sigma(A)} \quad & a.X = \mu + \epsilon_a \\
\forall_{c\in\sigma(C)} \quad & c.Z = \sum_{a\in ne^{\sigma}(c)\cap\sigma(A)} a.X + \mu + \epsilon_c \\
\forall_{b\in\sigma(B)} \quad & b.Y = \sum_{c\in ne^{\sigma}(b)\cap\sigma(C)} c.Z + \mu + \epsilon_b
\end{aligned}$$

and

$$\begin{aligned}
\forall_{c\in\sigma(C)} \quad & c.Z = \mu + \epsilon_c \\
\forall_{a\in\sigma(A)} \quad & a.X = \sum_{c\in ne^{\sigma}(a)\cap\sigma(C)} c.Z + \mu + \epsilon_a \\
\forall_{b\in\sigma(B)} \quad & b.Y = \sum_{c\in ne^{\sigma}(b)\cap\sigma(C)} c.Z + \mu + \epsilon_b.
\end{aligned}$$

For testing alternative hypothesis, we use the following model where (roughly speaking) Z is a common effect of X and Y,

$$\begin{aligned}
\forall_{a\in\sigma(A)} \quad & a.X = \mu + \epsilon_a \\
\forall_{b\in\sigma(B)} \quad & b.Y = \mu + \epsilon_b \\
\forall_{c\in\sigma(C)} \quad & c.Z = \sum_{a\in ne^{\sigma}(c)\cap\sigma(A)} a.X + \\
& \qquad \sum_{b\in ne^{\sigma}(c)\cap\sigma(B)} b.Y + \mu + \epsilon_c.
\end{aligned}$$

In all experiments, we set $\mu = 0.3$. We test

$$[B].Y \perp\!\!\!\perp [B, R_{AB}, A].X \mid [B, R_{BC}, C].Z$$

for the null hypothesis and test

$$[C, R_{AC}, A].X \perp\!\!\!\perp [C, R_{BC}, B].Y \mid [C].Z$$

for the alternative hypothesis.

C TYPE I ERRORS FOR DIFFERENT METHODS

We illustrate type-I error plots for Section 5.2 and 5.3.

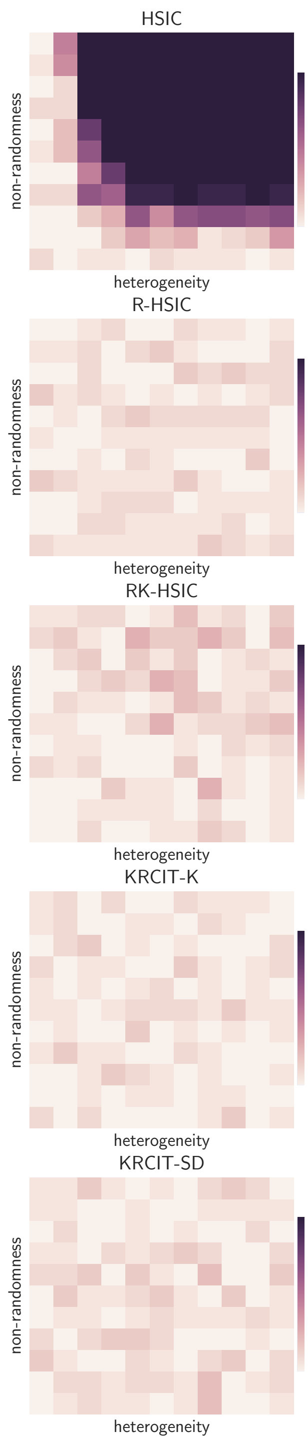

Figure 9: Type-I errors of different methods for Section 5.2.

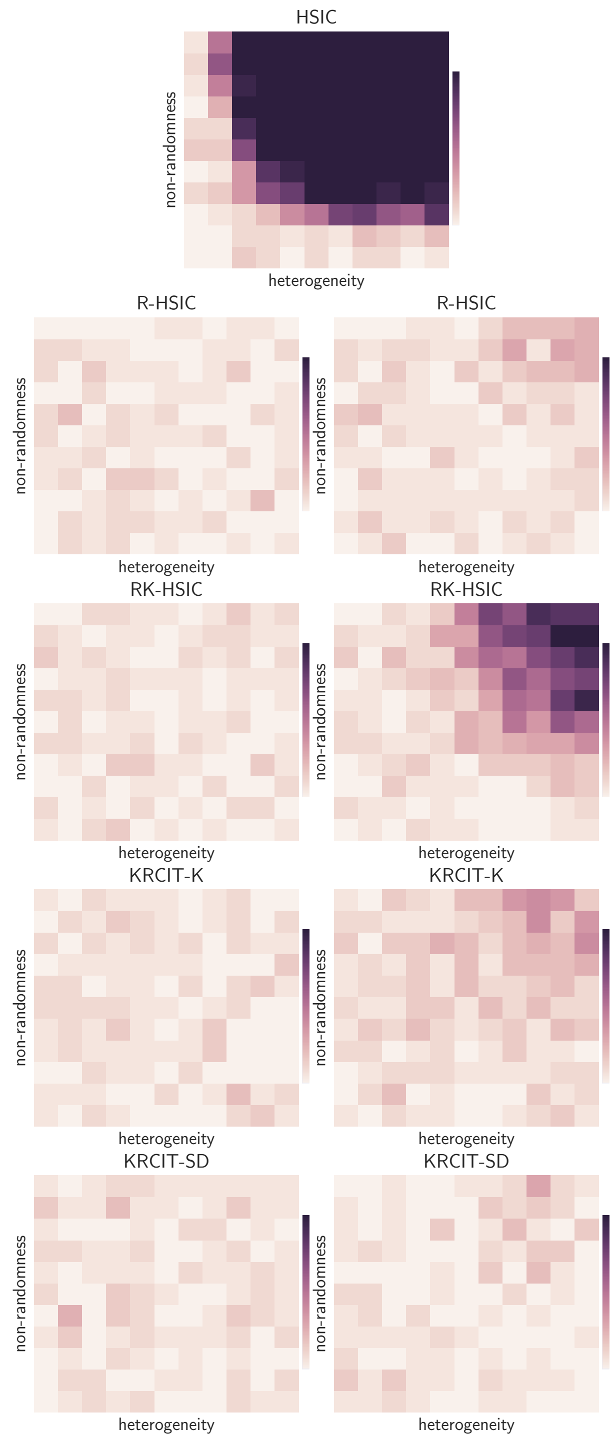

Figure 10: Type-I errors of different methods with two different contexts based on hop=1 (left) and hop=2 (right) for Section 5.3. HSIC does not use contexts.